



Early warnings of extreme winds using the ECMWF Extreme Forecast Index

Petroliagis, Thomas I.; Pinson, Pierre

Published in:
Meteorological Applications

Link to article, DOI:
[10.1002/met.1339](https://doi.org/10.1002/met.1339)

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Petroliagis, T. I., & Pinson, P. (2014). Early warnings of extreme winds using the ECMWF Extreme Forecast Index. *Meteorological Applications*, 21, 171–185. <https://doi.org/10.1002/met.1339>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Early warnings of extreme winds using the ECMWF Extreme Forecast Index

Thomas I. Petrolia^{a,*} and Pierre Pinson^{a,b}

^a European Centre for Medium-Range Weather Forecasts, Reading, UK

^b DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

ABSTRACT: The European FP7 SafeWind Project aims at developing research towards a European vision of wind power forecasting, which requires advanced meteorological support concerning extreme wind events. This study is focused mainly on early warnings of extreme winds in the early medium-range. Three synoptic stations (airports) of North Germany (Bremen, Hamburg and Hannover) were considered for the construction of time series of daily maximum wind speeds. All daily wind extremes were found to be linked to very intense surface cyclonic circulation systems being advected mainly by southwest and northwest flow regimes. Overall, it becomes clear that the first indications of an extreme wind event might come from the ECMWF deterministic and/or probabilistic components capturing very intense weather systems (possible windstorms) in the medium term. For early warnings, all available EPS Extreme Forecast Index (EFI) formulations were used, by linking daily maximum wind speeds to EFI values for different forecast horizons. From all possible EFI schemes deployed for issuing early warnings, the highest skill was found for the Gust Factor formulation (EFI-10FGI). Using EFI-10FGI, the corresponding 99% threshold could provide an early warning for a considerable portion of the >99% wind extremes, but not for all. By lowering this threshold the number of hits is increased until all extremes are captured (zero misses), although by doing so the number of false alarms increased significantly. Consequently, an optimal trade-off between hits and false alarms has to be made when setting different (critical) EFI thresholds.

KEY WORDS ensemble prediction system; predictability; rare severe events; early warnings; critical thresholds; alerts; alarm bells; gust factor

Received 23 January 2012; Revised 13 April 2012; Accepted 21 May 2012

1. Introduction

The Extreme Forecast Index (EFI) was developed at ECMWF as a tool to provide forecasters with general guidance on potential extreme events based on information from the Ensemble Prediction System (EPS). Verification results show that the EFI has substantial skill in providing early warnings of extreme events (Richardson *et al.*, 2011), confirming the subjective experience of forecasters in the Member States where the EFI is widely used.

The typical forecast horizon of the EFI has been the early medium-range. During this time interval, indications of an extreme weather event coming from the EFI are considered as ‘early warnings’. Beyond day 5, ‘alarm bells/signals’ also exist resulting from the ability of deterministic IFS (Integrated Forecasting System) and/or EPS components to capture very intense weather systems (possible windstorms) at medium- and late medium-range. Figure 1 contains a description of the different forecast and warning terms used in this study. Furthermore, this study considers the process by which forecasters can make full use of the EFI in order to provide local warnings of extreme events, with emphasis on warnings of high winds.

The concepts are illustrated by studying the extreme winds affecting three airports in Germany. Results of a synoptic study of extremes, skill assessment of the EFI and the setting of optimal EFI thresholds are presented. Finally, some examples of using the EFI are given. It is intended that the results

presented here will assist forecasters in making crucial decisions concerning severe weather.

2. Extreme events and predictability limitations

One of the most important tasks of National Meteorological Services is to help forewarn society about severe or high-impact events that can result in considerable damage and large losses (Casati *et al.*, 2008). Much of the benefit to society through improved weather forecasts will come from advances in our capability to forecast such events so that mitigating actions can be taken.

Severe events are usually considered to be rare, hence the use of the term ‘Rare Severe Event’ (RSE) by Murphy (1991). Such events are also loosely referred to as ‘Extreme Events’ in atmospheric science (WMO, 2011). Extreme events can come in many forms, such as intense multi-cell thunderstorms, tropical and extra-tropical cyclones, very intense wind events, heavy rain events, extreme heat and cold, floods and droughts.

Extreme events pose a special problem because they are infrequent, poorly documented by observations, and at the limit of predictability. Quantitative verification of such extremes is therefore more difficult and the statistical significance of verification results is mostly poor. At the same time, it is recognized that a poor numerical forecast in absolute terms can be of great value if it is well interpreted by an experienced forecaster. So, for example, the same absolute error may have various degrees of significance depending on how the forecast is placed with respect to climatology. The issue of extremes

*Correspondence to: T. I. Petrolia, ECMWF, Shinfield Park, Reading RG2 9AX, UK. E-mail: thomas.petrolia@ecmwf.int

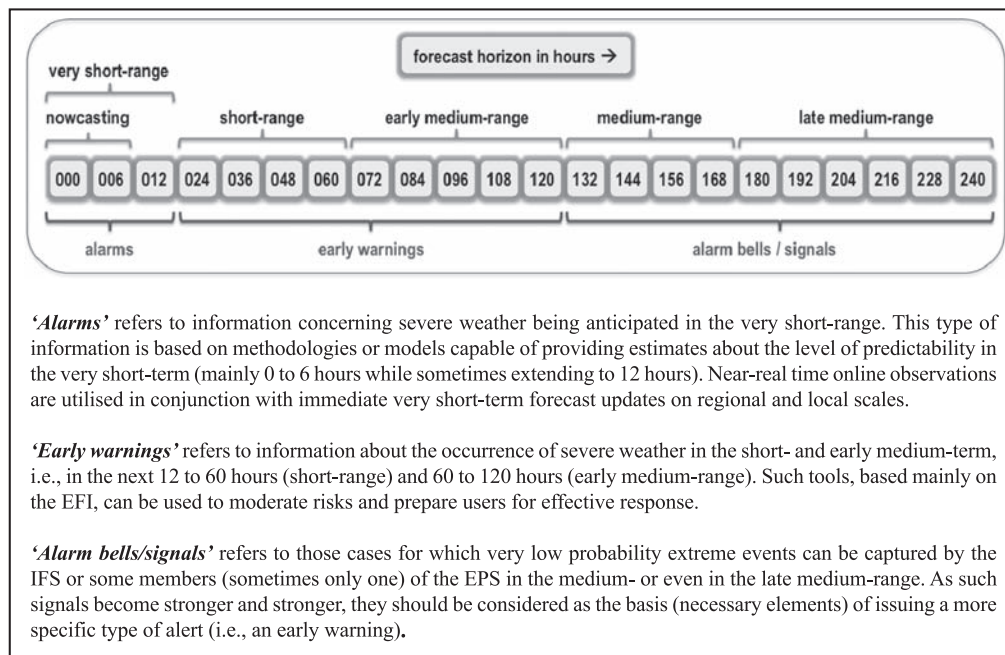


Figure 1. Description of different forecast and warning terms used in this study.

is made more complex by the scale difference between model and observations. In many cases one should not expect the current models to reproduce the maximum values of weather parameters observed in extreme events because their resolution is relatively low. However, methods should be designed to diagnose severe weather based on the existing models, and the validity of these diagnostics should be thoroughly verified (Bougeault, 2003).

In operational forecasting, a 'gap' seems to exist between the events for which forecasters need to issue early warnings and what the numerical model guidance can provide. Some types of weather responsible for damage (e.g. lightning, wind gusts at different heights and fog) might not be explicitly simulated (predicted) by the model, and must therefore be diagnosed from other variables. Even if a type of weather can be explicitly predicted (e.g., heavy rain), the model resolution might not capture its intensity; this could be because the processes associated with the variable are often at sub-grid scale. Some mesoscale models are being run experimentally at resolutions of 1–2 km, but most operational mesoscale models have grid scales of 5–15 km, and global models are even coarser.

A study of past extreme wind events (such as windstorms) reveals that only a small proportion of ensemble members (or of deterministic forecasts from different weather centres) succeeded in predicting severe storms, even about 24 h in advance (Legg and Mylne, 2004). Furthermore, in a synoptic situation where severe weather is possible, once a forecast moves into the chaotic non-linear regime most ensemble members are likely to be drawn towards the model's climatology. This seems to be the case (with the central control forecast predicting severe weather and perturbed analyses leading to less severe conditions), even when it is one or more perturbed ensemble members that predict severe conditions. Based on this, the forecast probability density function is always likely to be skewed away from severe weather. Thus, although the ensemble can be expected to include members with severe events, it would be unusual for it to predict (i.e. to assign) high probabilities of those events.

Since the above analysis applies equally well to the real atmosphere as to a model describing the atmosphere, it can be argued that the occurrence of severe weather is fundamentally a low probability event in the atmosphere. Thus, on most occasions it should only be appropriate to issue early warnings with a low probability, since early warnings and/or alarm bells/signals communicated to end users a few days ahead of potential events are of significant benefit.

3. Extreme events and the EFI

One of the aims of ensemble prediction is to improve the forecasting of severe weather. To the extent that the development of severe weather is frequently highly non-linear and therefore sensitive to forecast errors, this seems as an appropriate application of ensembles. On the other hand, the ability of current models to generate extreme/severe storms has improved in recent years. Based on such models' abilities, issuing early warnings of extreme events seems possible. However, what constitutes an extreme event depends on location or season. To quantify this notion of extreme weather events the Extreme Forecast Index (EFI) has been developed (Lalauette, 2003).

The EFI measures the difference between the probability distribution from the EPS and the model climate distribution. The underlying assumption is that if a forecast is extreme relative to the model climate, the real weather is also likely to be extreme compared to the real climate. The EFI is formulated so that it lies between -1 and $+1$. This index seems capable of revealing whether the deviation from climate is in a direction that may be dangerous for human activity (Lalauette, 2003). If the EFI indicates a potential severe weather event, the forecaster can then examine more detailed information from the forecast to make a more thorough assessment of the risk to the end users. Since the model climate accounts for the variability of the weather parameters in both geographical location and time of year, the EFI allows the user to identify an anomalous weather situation without having to define specific thresholds for an extreme event.

On the other hand, probabilistic forecasts provide useful information on the uncertainty: it is desirable to communicate such information for events that can incur large losses. Probabilistic forecasts can also be used to assess risk quantitatively using a cost-loss model (for example) and help in optimal decision-making for specific users. It should be noted that signal detection of severe weather events in the medium-range is likely to be difficult. Bearing this in mind, issuing 'alarm bells/signals' at medium- and late medium-range based on predicting intense cyclonic circulation systems and/or synoptic weather types linked to destructive wind events should not happen automatically. Rather these 'alarm bells/signals' should act as a 'warning light' that ensures a potentially dangerous event does not go unnoticed by the forecasters and end users.

In terms of the EFI, emphasis has been put on assessing the predictability of severe weather in the short- and early medium-range as a basis for issuing early warnings. However, it has been shown that case studies provide biased estimators of severe weather forecast performance: there is always some kind of signal that the forecaster should, in retrospect, have been aware of. In contrast, conducting a verification study using forecasters' expertise in real time or delayed mode is both costly in terms of human resources and biased in its own way by the forecasters' perspective (Lalaurette and Grijn, 2005).

The events targeted here have been those daily wind speed extremes that exceed the 99th percentile of the model and station (synoptic) climate records. It will become clear that the corresponding 99% EFI threshold, which can substantiate the issuing of an early warning, provides useful guidance for a considerable portion of >99% wind extremes. Note that during the period covered by this study, December 2003 to May 2010, the resolution of the EPS has changed. Up to February 2006 it had a horizontal resolution of TL255L40 (~80 km), thereafter, to January 2010, it had a resolution of T399L62 (~50 km) out to 10 days; it then increased to T639L62 (~30 km). Besides the EPS, which is capable of providing information on uncertainty, the IFS platform can provide useful deterministic forecast guidance about the weather for many different end users and the general public. The deterministic IFS (Simmons *et al.*, 1988) currently has a horizontal resolution of 16 km (T1279) and 91 levels in the vertical, while the EPS has a horizontal resolution of 32 km and a vertical resolution of 62 levels. The EPS was implemented operationally 20 years ago (Palmer *et al.*, 1993; Molteni *et al.*, 1996) and has undergone many changes since then (Palmer *et al.*, 2007).

In practical terms, the value of an ensemble prediction system is that it gives forecasters the means to access quantitatively the risk of weather sensitive events occurring in the days ahead. The current EPS comprises 50 + 1 members. The probability of a given event is determined from the fraction of ensemble members, which predict the event. Verification of probability forecasts requires many matched forecasts and observations. This may be difficult to achieve for high impact weather, which is often rare by definition. Verification using only a small dataset leads to results with large uncertainties.

4. Case study for Bremen, Hamburg and Hannover airports

The main methodology of linking wind extreme events to the EFI has been investigated for three synoptic stations (airports) in North Germany: Bremen, Hamburg and Hannover (shown in Figure 2). Results concerning the possibility of issuing 'early warnings' in the short- and early medium-range to users

are documented. Furthermore, the synoptic circulation patterns linked to extremes are examined. The possibility of providing additional critical information, such as 'alarm bells/signals', in the medium- and late medium-range of potential windstorms is also investigated. Note that the distinction between early warnings and alarm bells/signals terms follow the definitions given in Figure 1. The former refers to the short- and early medium-range (12–120 h) while the latter to the medium- and late medium-range (120–240 h) forecast horizons.

4.1. Synoptic investigation of extremes

For the definition of wind speed extremes, the ECMWF ERA-Interim climatological database (Simmons *et al.*, 2007) has been used. A time series of daily maximum wind speeds was constructed for each station, spanning 2374 days (from 1 December 2003 to 31 May 2010). These values represent the reanalysis daily wind speed maximum ('Reanalysis' mode) defined as the maximum value of the four plus one (the 0000 value of the next day) synoptic-hour values spanning the 24 h interval of each day (i.e. 0000, 0600, 1200 and 1800 UTC values). Similarly, a time series was constructed based on each station's observations of maximum wind speed ('Observation' mode). The difference here is that daily maximum values are defined by considering all eight reported observations at 0000, 0300, 0600, 0900, 1200, 1500, 1800 and 2100 UTC (i.e. maximum value over the day, defined as the interval between 0000 and 2359 UTC).

Expressing daily reanalysis maximum values as anomalies (from the ensemble mean), a specific type of daily maximum anomaly time series is constructed for each station in both 'Reanalysis' and 'Observation' modes. For each station and for all extremes belonging to the >99% category, the synoptic meteorological environment was investigated. All extremes were found to be linked to very intense circulation systems (surface pressure lows) affecting all three stations during the same day most of the times, as clearly shown in Table 1. The pronounced relationship between extremes belonging to



Figure 2. Geographical position of Bremen, Hamburg and Hannover airports/synoptic stations in North Germany (denoted by white circles).

Table 1. Dates and names of intense surface lows linked to >99% daily extremes in 'Reanalysis' mode for Bremen, Hamburg and Hannover airports/stations.

Date	Surface low identifier	Bremen	Hamburg	Hannover
21 December 2003	Jan	*	—	*
13 January 2004	Hanne	*	*	*
14 January 2004		*	*	*
31 January 2004	Pia and Quinne	*	*	*
1 February 2004		*	*	*
20 March 2004	Melita and Nina	*	*	*
1 March 2004	Oralie and Paloma	*	*	*
17 November 2004	Pia (New)	—	*	—
18 November 2004		—	*	—
2 January 2005	Alloys	*	—	—
8 January 2005	Dimitri and Erwin	*	*	*
12 February 2005	Ulf	*	*	*
17 March 2005	Heijo and Iradj	*	*	*
30 December 2006	Karla and Lotte	*	*	*
31 December 2006		*	*	*
11 January 2007	Franz and Anonym	*	*	*
12 January 2007	Gerhard and Hanno	—	*	—
13 January 2007		—	*	—
18 January 2007	Kyrill	*	*	*
19 January 2007	Kyrill and Lancelot	*	—	*
21 January 2007	Lancelot	—	—	*
10 April 2007	Xenophon	—	*	—
11 May 2007	Ewald I and II	—	—	*
26 June 2007	Uriah and Vanni	—	—	*
27 June 2007	—	—	*	—
26 January 2008	Paula	—	*	—
31 January 2008	Resi	*	*	*
1 February 2008		*	*	—
1 March 2008	Emma	*	*	*
2 March 2008	—	*	—	*
12 March 2008	Johanna and Kirsten	*	—	*
23 March 2009	Herbert	*	—	*
3 October 2009	Ralf and Soeren	*	*	—
16 October 2009	Vimar and Xavier	*	—	—
18 November 2009	Ingmar and Jorgen	*	*	*
1 March 2010	Xynthia	—	—	*

An asterisk is used to denote events belonging in the >99% extreme category.

the >99% category and very intense cyclonic systems could constitute the basis of issuing 'alarm bells/signals'. This seems possible since both the deterministic IFS and EPS are capable of generating severe storms in the medium- and even in the late medium-range.

4.2. Use of the DWD Objective Weather Type Classification

Further investigation of the synoptic situation associated with the extremes has been performed by examining all aspects of relationships between the large-scale atmospheric circulation on one side and surface climate and environmental variables on the other. The Objective Weather Type Classification (OWTC) methodology of the National German Weather Service (DWD)

Table 2. Frequencies of daily wind extremes belonging to the >99% percentile category for various advection flow regimes.

Basic weather type advection at 700 hPa	Frequency (%)
SW (southwest)	50
NW (northwest)	50
NE (northeast)	0
SE (southeast)	0
No prevailing advection	0

(Bissolli and Dittmann, 2001) uses meteorological criteria such as:

- 700 hPa advection ('No advection', 'Northeast', 'Southeast', 'Southwest' and 'Northeast');
- Cyclonicity 950 hPa ('Cyclonic', 'Anticyclonic');
- Cyclonicity 500 hPa ('Cyclonic', 'Anticyclonic'), and,
- Humidity from 950 to 300 hPa ('Wet', 'Dry').

These lead to numerical indices from which the weather types are derived. This is an objective procedure that is defined unambiguously. Using all OWCT criteria, 40 weather types are derived. In contrast to the widely used 'Grosswetterlagen' of Hess and Brezowsky (Hess and Brezowsky, 1977; Nicolis *et al.*, 1997), which are determined by a subjective method, OWCT's objective procedure is numerically reproducible at any time with the same classification result. Here, for simplicity, the categorization is performed using only one criterion: the advection at 700 hPa. All weather types that prevailed over North Germany are considered for the same interval of 2374 days. In this way a weather type time series, harmonized to reanalysis and observation daily maximum wind anomalies, is constructed. In simple words this means that for every daily maximum anomaly (in 'Reanalysis' or 'Observation' mode) a distinct weather type is assigned.

Studying closely the elements of Table 1, it became clear that all >99% extremes were associated with very intense surface lows being advected in either southwest or northwest regimes, with 50% falling into each category (Table 2). None of the extremes was associated to northeast or southeast regimes or the no advection category (i.e. no prevailing advection or no advection). These results seem to agree quite well with those by Donat (2010) who found that about 80% of storms affecting Central Europe are associated with westerly flow regimes.

For operational forecasting, these results suggest that if the advection of an anticipated intense cyclonic system falls into one of the critical southwest or northwest regimes special attention should be given. Nevertheless, extra detail about the time, place or intensity of the event can be added nearer the time of the event. On the other hand, issuing early warnings for the short- and early medium-range should be based on more concrete (objective) criteria. A probabilistic approach is one way for extracting a useful signal from numerical forecasts in a way that can be tailored to the specific needs of users, so the decision of issuing an early warning should be based on the intensity (i.e. using a critical threshold) of the EFI.

5. The possibility of issuing early warnings based on the EFI

The EFI is not only sensitive to a shift in the tails (i.e. in the extremes) but also in the median of the forecast distribution.

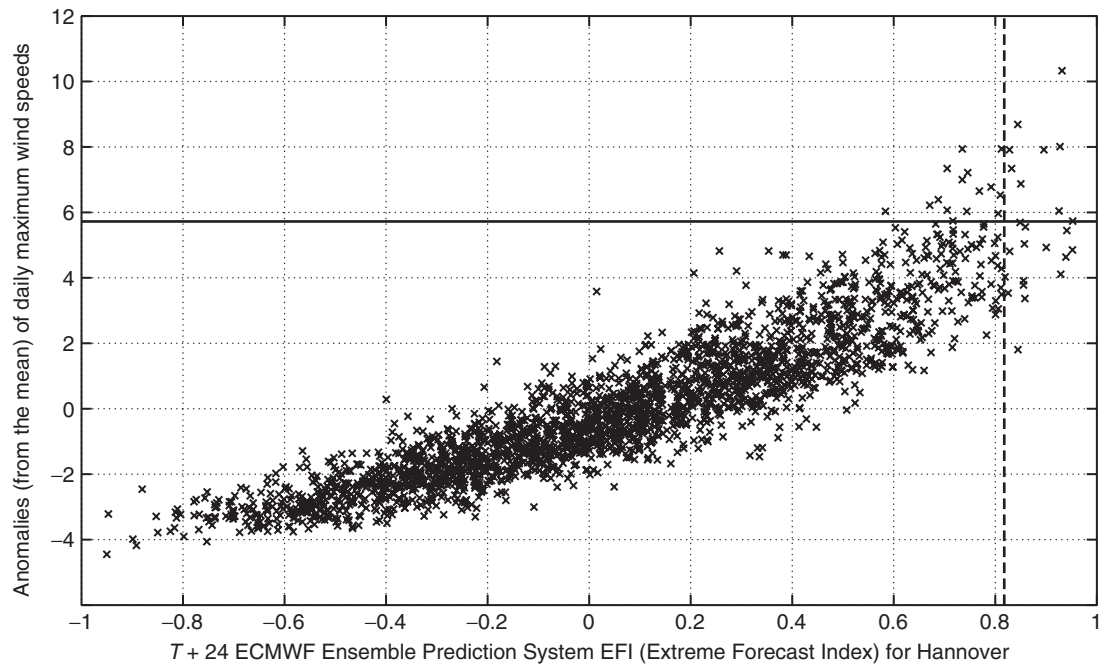


Figure 3. Example of anomalies of daily maximum 10 m wind speeds in 'Reanalysis' mode against 24 h forecasts of EFI-10FGI (based on 0000 UTC) values for Hannover. The dashed vertical line represents the 99% EFI threshold, while the solid horizontal line is the 99% percentile value of maximum daily wind speed anomalies.

In other words, high values of EFI might be achieved either because there is a limited number of members showing extreme values with respect to climate or, for example, because almost all members are showing only a moderate departure from the climate. The EFI values are, therefore, also a function of the EPS spread. This means that small EPS spread facilitates EFI extremes. Nevertheless, EFI still represents a very useful tool that easily allows the identification of extremes with respect to location and season. A North European forecaster, who in spring sees the EFI warning about extremely low temperatures in the Mediterranean area, should, of course, realize that the weather might not be extremely cold from a Scandinavian point of view. One should bear in mind that EFI values cannot replace probabilities: it just put them into perspective. It should be stressed once more that the EFI is a parameter giving an early warning to forecasters and end users.

In the present study two formulations of the EFI were used: 10FGI (based on a maximum wind gust) and 10WSI (based on instantaneous 10 m wind). For each formulation all sets of EFI forecasts based on both initialization times (i.e. 0000 and 1200 UTC) were considered in 'Reanalysis' and 'Observation' modes. Clear signs that EFI values are closely linked to daily maximum wind speeds are contained in Figure 3. The $T + 24$ step is being used as a demonstration example here, but similar results apply for the rest of the forecast horizons. These results reveal beyond any doubt that all reanalysis daily extremes (falling in the >99th percentile category) for Hannover correspond to strong positive EFI-10FGI values (based on 0000 UTC runs). Furthermore, the 99th EFI percentile threshold seems able to provide an early warning for a considerable portion of >99% wind extremes (hits).

6. Skill assessment of the EFI

Besides EFI's skill assessment over selected points, e.g. Bremen, Hamburg and Hannover, resembling single wind farm

environments, average values of wind maxima for all three stations and corresponding (averaged) values of EFI were also taken into consideration. This set of averaged values (named as BHH Area) resembles situations prevailing over a greater area that contains a number of wind farms for obvious upscaling purposes.

Results in terms of hit rates and false alarm rates for different EFI thresholds values are studied by using ROC diagrams and more specifically ROCA (Area under the ROC Curve) values. In terms of ROCA, the EFI-10FGI gust factors based on 0000 UTC and 1200 UTC are comparable in skill in 'Reanalysis' mode, both comprising high (skilful) values, with the 0000 UTC formulation slightly more skilful. Furthermore, EFI forecast guidance over single points seems to be equally skilful over the greater area containing all three points (BHH Area). For EFI-10WSI no significant difference in skill was detected between forecasts based on 0000 UTC and 1200 UTC in 'Reanalysis' mode. Also, skill values of EFI-10WSI for selected points were found to be comparable to those obtained over the BHH Area.

As in the 'Reanalysis' mode, in the 'Observation' mode there were no significant differences between using 0000 and 1200 UTC data in EFI-10FGI. The same applied to EFI-10WSI. However, for both the EFI-10FGI and EFI-10WSI formulations the forecasts are found to be less skilful in the 'Observation' mode. This is not surprising because the EFI uses its own model climate and not the real climate (climatology). Another obvious reason might be that the model has an easier task verifying against its own analysis (reanalysis for this case) extremes than against real (synoptic) observational extremes.

Overall, EFI-10WSI was found to be less skilful than EFI-10FGI. This could be anticipated since a daily series of extreme wind values was constructed that are different to the mean (daily averaged) wind time series in both the 'Reanalysis' and 'Observation' modes. Going after such extremes, the EFI-10FGI formulation being based on model's gust factor ('gusty')

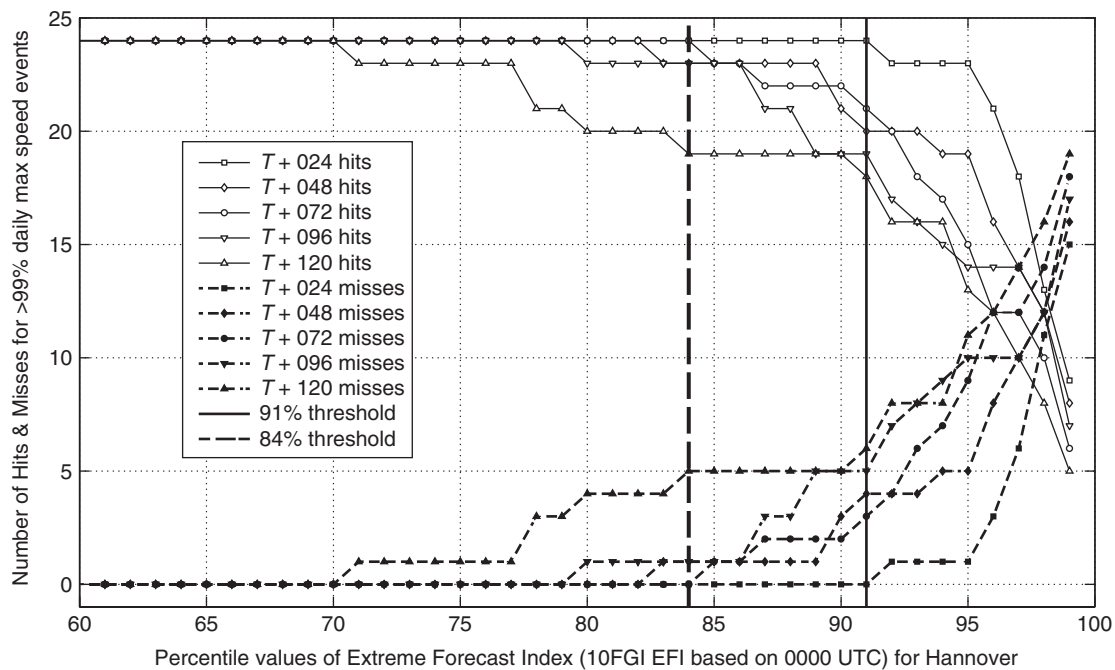


Figure 4. Hits and misses for the >99% category wind extremes based on different EFI-10FGI (0000 UTC) thresholds for various forecast horizons (Hannover). The 91% EFI threshold (resulting to ‘zero misses’ for day 1) and the 84% threshold (‘zero misses’ for day 3) are plotted as well (solid and dashed vertical lines respectively).

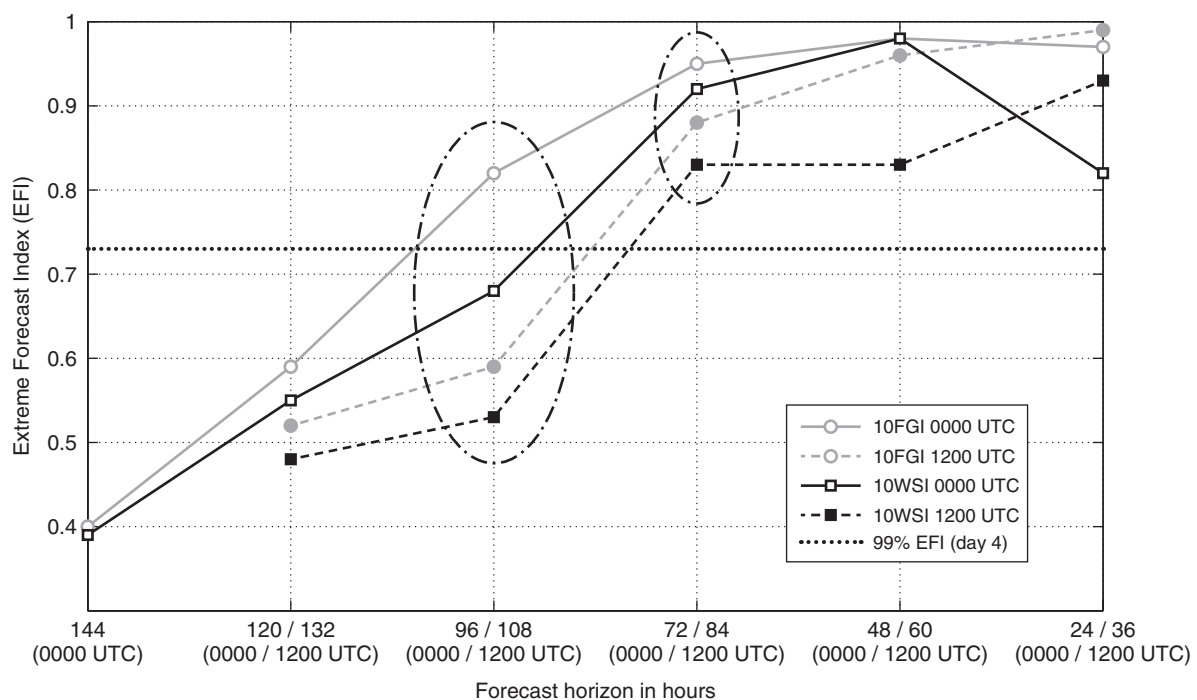


Figure 5. EFI-10FGI and EFI-WSI (based on 0000 and 1200 UTC) values for Xynthia's maximum impact area located at the borders of Luxembourg and France (28 February 2010).

components, seems a more appropriate option than the EFI-10WSI formulation that is based on ‘normal’ instantaneous 10 m wind components. Nevertheless, results in predicting extremes by using EFI indicate beyond any doubt significant skill in both the short- and early medium-range. However, it should be pointed out that to achieve large hit rates for all forecast horizons (as in the example shown in Figure 4) a significant number of false alarms would be generated as well.

This behaviour is somewhat hidden by the rarity of the rare severe events represented in ROC curves and the associated ROCA scores (Choo, 2009). The rate of hits/false alarms generated by using such an early warning platform would clearly be very far from what can be obtained by waiting until a few hours before the event, but because there are protection measures that users have to take in advance, and they cannot wait until the last minute to be implemented, there is a certain

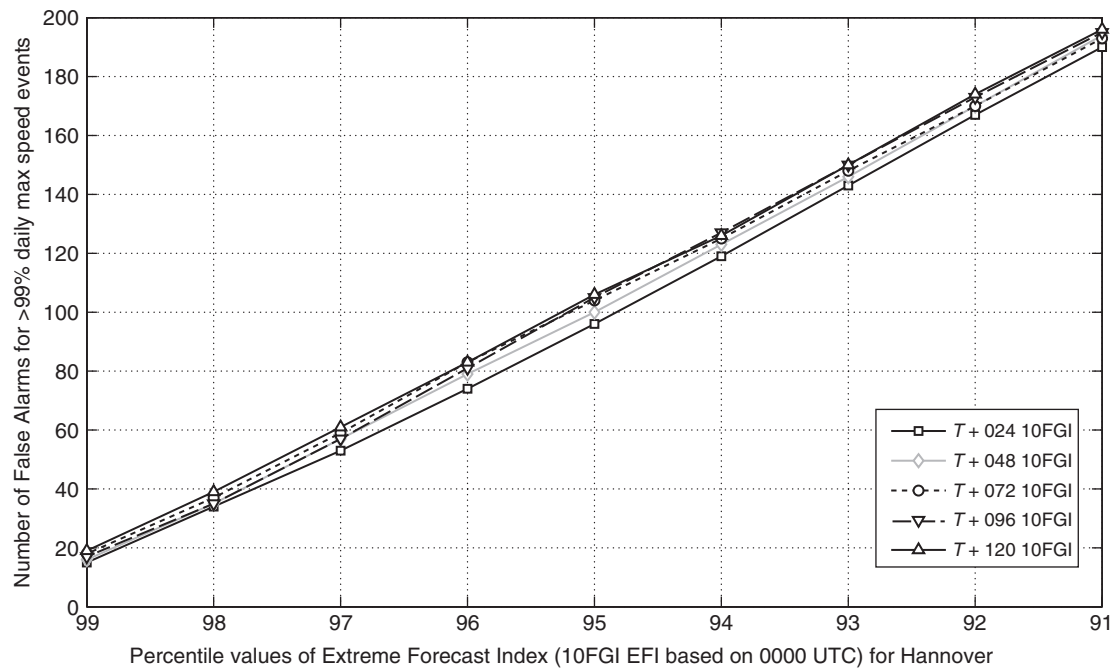


Figure 6. Number of false alarms for different EFI thresholds for horizons spanning from 24 to 120 h (Hannover). It is obvious that the $T + 24$ 'zero misses' threshold (91%) introduces 190 false alarms.

value attached to early warning procedures even if they relate to a significant number of false alarms.

7. Setting an optimal EFI threshold

The usefulness of an early warning based on the EFI can be seen in Figure 5. This shows the EFI formulations for the maximum impact position (borders of Luxembourg and France) of storm Xynthia on 28 February 2010 (Meteo-France, 2010). It is clear that the EFI-10FGI based on 0000 UTC is capable of providing an early warning 4 days in advance, since its value (0.82) is found to be higher than the 99% EFI percentile threshold (being equal to 0.73). The same holds for the rest of EFI formulations but there is a certain delay of about 24 h.

Using the 99% EFI percentile threshold, very high (skilful) ROCA values were found for all three airports used in the study. It is important to point out, though, that the 99% threshold is capable of providing an early warning for a certain portion of extremes, but not for all (as displayed in Figure 4). By lowering this threshold, the number of hits is increased until all extremes are captured, but the number of false alarms is increased significantly. This unavoidable drawback can be seen in Figure 6, where the number of false alarms is plotted against different EFI-10FGI thresholds for Hannover airport corresponding to the data (hits) contained in Table 3. Furthermore, by using the 99% EFI threshold the number of misses (of >99% extremes) equals the number of false alarms for all forecast horizons. In a perfect forecasting environment this common number (of misses and false alarms) would be equal to zero, so such a system would provide only hits (correct forecasts) and correct negatives (non-events). In reality, however, misses and false alarms do exist.

A clear demonstration of this unavoidable limitation can be seen in Table 3. The number of hits for the 24 h forecast is equal to 9, but there are also 15 misses and 15 false alarms. This means that the 99% EFI percentile threshold

Table 3. Number of hits for >99% category extremes based on various EFI-10FGI (0000 UTC) thresholds for different forecast horizons valid for Hannover (maximum number of hits: 24).

EFI threshold (%)	Day1 $T + 24$	Day2 $T + 48$	Day3 $T + 72$	Day4 $T + 96$	Day5 $T + 120$
70	24	24	24	24	24
71	24	24	24	24	23
72	24	24	24	24	23
73	24	24	24	24	23
74	24	24	24	24	23
75	24	24	24	24	23
76	24	24	24	24	23
77	24	24	24	24	23
78	24	24	24	24	21
79	24	24	24	24	21
80	24	24	24	23	20
81	24	24	24	23	20
82	24	24	24	23	20
83	24	23	24	23	20
84	24	23	24	23	19
85	24	23	23	23	19
86	24	23	23	23	19
87	24	23	22	21	19
88	24	23	22	21	19
89	24	23	22	19	19
90	24	21	22	19	19
91	24	20	21	19	18
92	23	20	20	17	16
93	23	20	18	16	16
94	23	19	17	15	16
95	23	19	15	14	13
96	21	16	12	14	12
97	18	14	12	14	10
98	13	12	10	12	8
99	9	8	6	7	5

The 'zero misses' EFI threshold for different forecast horizons is denoted by a black cell.

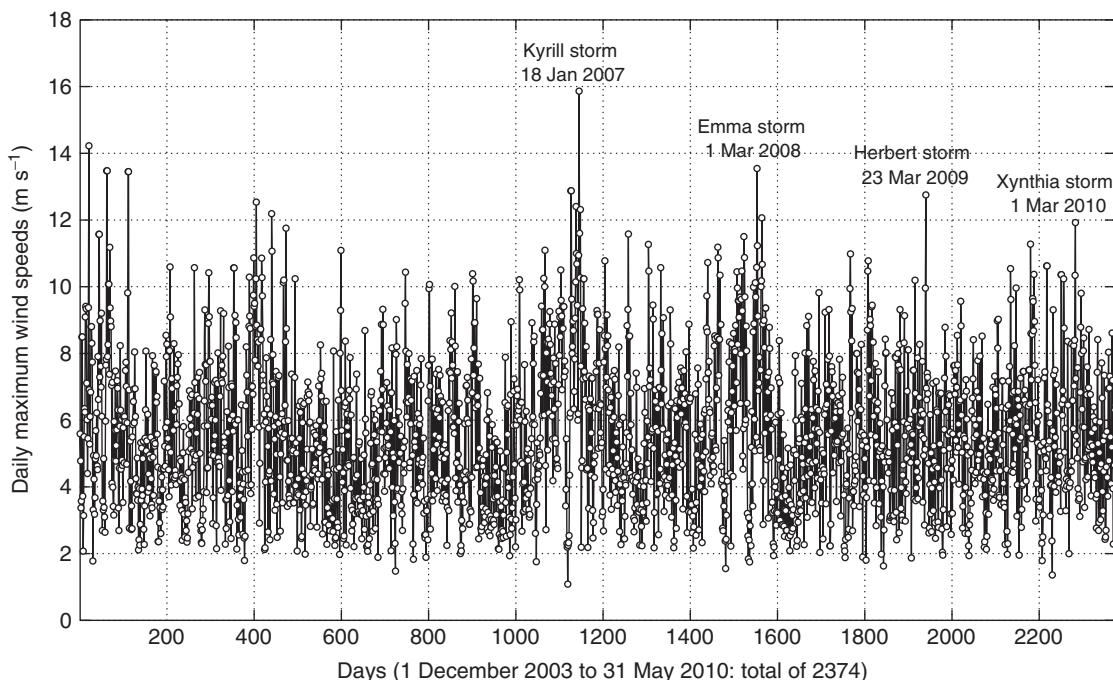


Figure 7. Time series of daily maximum wind speed values for Hannover over a period of 2374 days (1 December 2003 to 31 May 2010) in 'Reanalysis' mode. Peak values corresponding to Kyrill, Emma, Herbert and Xynthia storms are highlighted.

is able to provide an early warning for a portion of $>99\%$ wind extremes but not for all. By lowering this threshold the number of hits is increased (as shown in Figure 4) until all extremes are captured. The 'zero misses' EFI threshold (i.e. the one corresponding to the 91st percentile) highlighted by grey shading is able to provide the maximum of the total 24 hits (i.e. zero misses), although by doing so the number of false alarms is increased significantly and reaches 190. This limitation becomes more pronounced when different (longer) time horizons are to be considered, as easily seen by examining the different columns of Table 3. For instance, the day 5 'zero misses' for the 99th percentile extreme wind anomalies corresponds to a considerable lower threshold of EFI, equal to the 70th quantile (Wilks, 2006), resulting in 688 false alarms. Furthermore, a close examination of the cases linked to false alarms was performed relative to prevailing flow regimes, but no significant relationship could be established.

Overall, it is clear that all extremes (falling in the $>99\%$ category) are linked to strong positive EFI values. From all possible EFI schemes, the highest skill in issuing early warnings is linked to the EFI-10FGI formulation. This means that using EFI-10FGI (based on 0000 UTC), the corresponding 99th percentile threshold can provide an early warning for a certain portion of the $>99\%$ extreme category, but not for all. By lowering this threshold, the number of hits is increased till all extremes are captured although by doing so, the number of false alarms is increased significantly.

8. Examples of using the EFI

8.1. Case studies

The issue of setting optimal EFI thresholds is further investigated for extreme events over Hannover. All daily maximum wind speed values for Hannover ('Reanalysis' mode) over a period of 2374 days are plotted in Figure 7. A selection of

the four most recent spikes has been made (corresponding to a higher resolution EPS). These spikes indicate the following storms: Kyrill (18 January 2007), Emma (1 March 2008), Herbert (23 March 2009) and Xynthia (1 March 2010). An obvious reason for such a selection is the better model resolution available in more recent years.

As an example, the different EFI (10FGI) maps valid for storm Kyrill (Fink *et al.*, 2009) are displayed in Figure 8 for various forecast horizons: $T + 132$ (a), $T + 96$ (b) and $T + 48$ (c). Both the 95 and 98% EFI thresholds (highlighted by dotted lines), used here as critical thresholds, are able to provide an early warning for the windstorm from day 5.5 onwards as clearly seen in the relevant EFI-GRAM (d). Kyrill caused widespread damage across Western Europe, especially in the UK and Germany. There were 47 fatalities reported, as well as extensive disruptions of public transport, power outages to over 100 000 homes, severe damage to public and private buildings and major forest damage. An estimation of the insurance market loss was about 3.5 billion Euros. It is worth mentioning that it is quite uncommon for winter storms with such intensity as the one of Kyrill to reach Central and Eastern Europe.

Another example, for storm Emma, is displayed in Figure 9 for the same forecast period as in the Kyrill case. Once more it becomes clear that both the 95 and 98% EFI thresholds (highlighted by dotted lines) are capable of providing an early warning for the windstorm from day 5.5 ($T + 132$ h) onwards. Emma was a severe extratropical cyclone that was felt by many Europeans, with its most devastating impact on 1 March 2008, leaving behind at least 12 people dead in Austria, Germany, Poland and the Czech Republic. Wind speeds reached up to 166 km h^{-1} in Austria, and up to 180 km h^{-1} in Germany. Major infrastructure disruptions and some injuries were also reported in Belgium, France, Switzerland and the Netherlands. In Germany, gale-force winds toppled power lines which knocked out more than 5000 transformer stations across the country, cutting power to hundreds of thousands of homes. The strong winds closed roads and railways, overturned cars

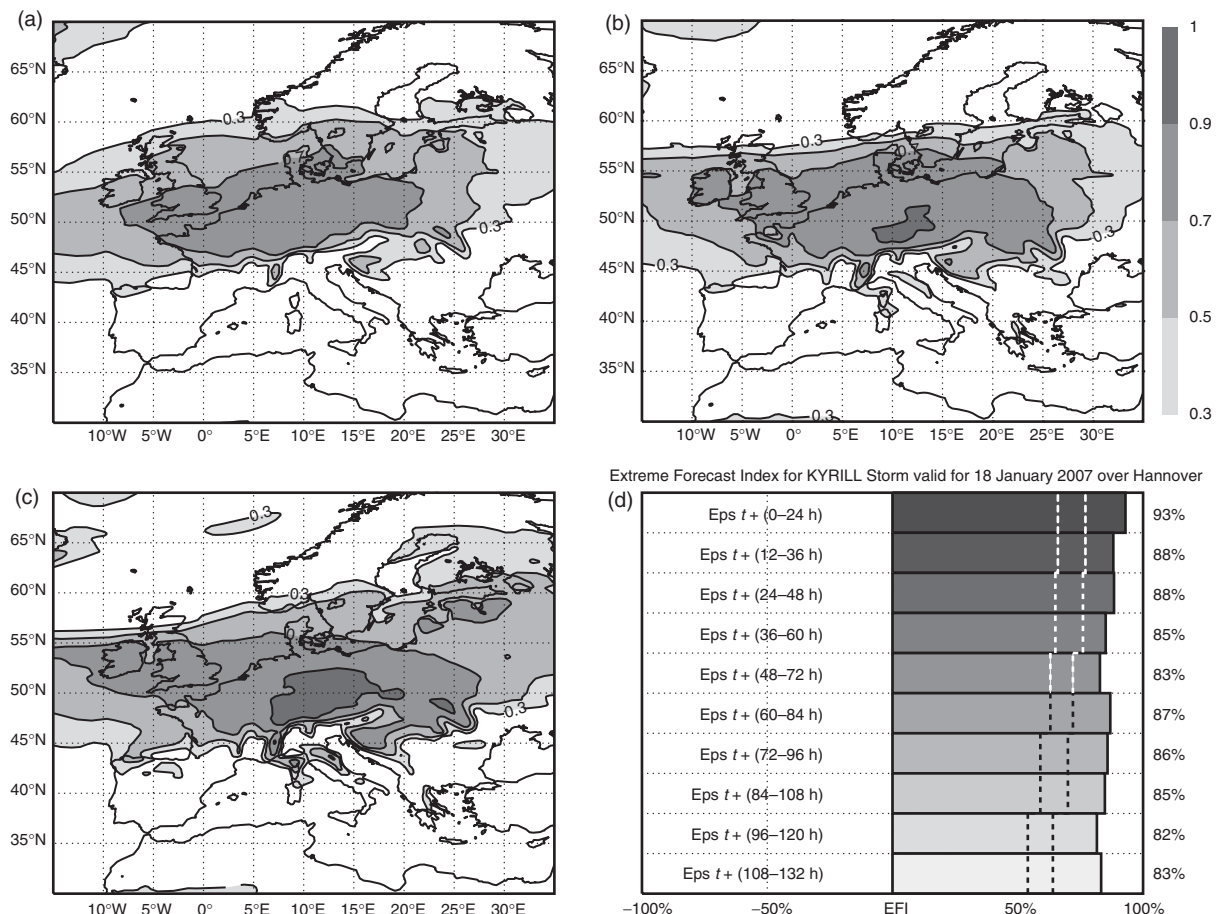


Figure 8. Samples of different EFI (10FGI) maps valid for the Kyrill storm hitting Hannover airport on 18 January 2007. Various forecast horizons are shown here: $T + 132$ (a), $T + 96$ (b) and $T + 48$ (c). A set of such maps is used in operational mode for the production of specialized 'EFI-GRAM' products as the one contained in panel d (valid for Hannover area).

and caused damage to several properties. According to reports, the southern state of Bavaria was particularly badly hit, where electricity was cut to 150 000 homes and where heavy rain caused flooding. Transport was disrupted, with reports of the violent winds causing the cancellation of nearly 200 flights from Frankfurt airport (Carpenter, 2008).

In order to investigate if the 95 and 98% percentiles can provide early warnings for the other storms considered here, Figure 10 is constructed. It has been obvious that both EFI thresholds work quite well for the storms Kyrill and Emma, but they seem to be inadequate for Herbert and Xynthia. More specifically, for Herbert, the 98% threshold fails to forewarn the user, while the 95% percentile seems to do a better job for horizons shorter than 84 h. As for Xynthia, the 98% threshold seems to work only for the 96 h horizon, while the 95% threshold works for all horizons shorter than 120 h (except for the 36 h one). As already pointed out, by lowering these thresholds the number of hits (for Herbert and Xynthia for example) will be increased, but by doing so the number of false alarms is also increased. Nevertheless, for both Herbert and Xynthia, a slightly lower threshold (say a value between 90 and 95%) could have resulted in users being warned quite in advance (i.e., more efficiently) about the potential impact of the approaching storms. That is why high (positive) values of EFI even lower than critical threshold(s) should always act as a 'red warning light' resulting to further investigation by the forecaster on the bench.

8.2. Operational interactive (clickable) EFI

In real-time operational mode, the current interactive (clickable) EFI, known as I-EFI (Figure 11) can be used to identify areas where the ensemble forecast distribution is significantly different from the climatological distribution in a global scale, and visualize any grid point distributions (zooming over points of interest). An example of I-EFI capabilities for the extreme winds that prevailed on 1 December 2011 in the west areas of USA is presented in Figure 11. The destructive gusts were produced by two separate weather systems that channelled cold air from the north into California, Nevada, Utah and Colorado states. Winds gusts over 100 mph were recorded locally, while more than 380 000 homes lost power. Thousands of trees snapped, blocking roads and damaging property. Scores of schools were closed and motorists battled gridlock caused by broken traffic signals and blowing debris. The storm, which produced some of the strongest wind gusts in more than a decade, was caused by a highly unusual weather system as a strong anticyclone moved over the Pacific Northwest beneath a large upper tropospheric ridge. Low pressure over the southwestern United States combined with this anticyclone to the north to produce high impact strong easterly flow over the western U.S..

Using I-EFI's capability to zoom over preselected areas of interest, zooming over North America is possible. Such a magnified and more detailed I-EFI map is presented in Figure 12 comprising the anomalous (and possibly extreme)

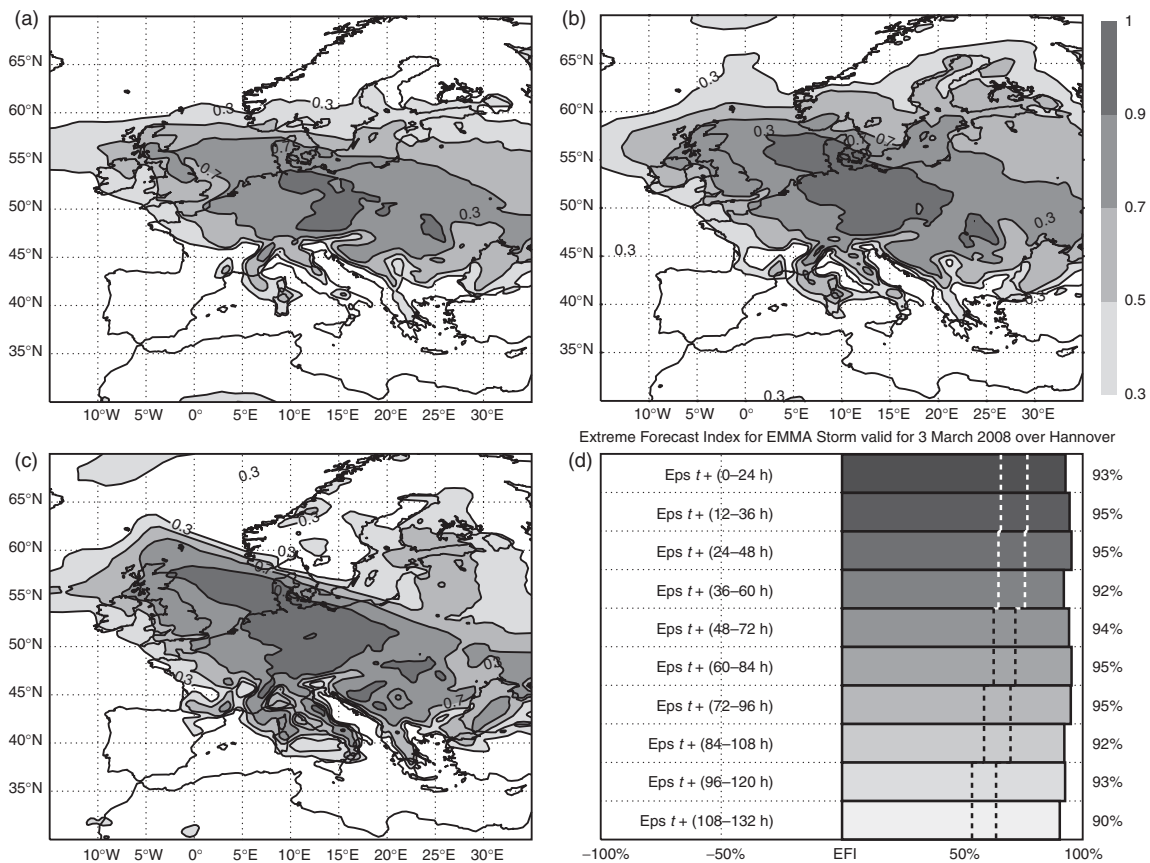


Figure 9. As in Figure 8, but for the Emma storm hitting Hannover airport on 1 March 2008.

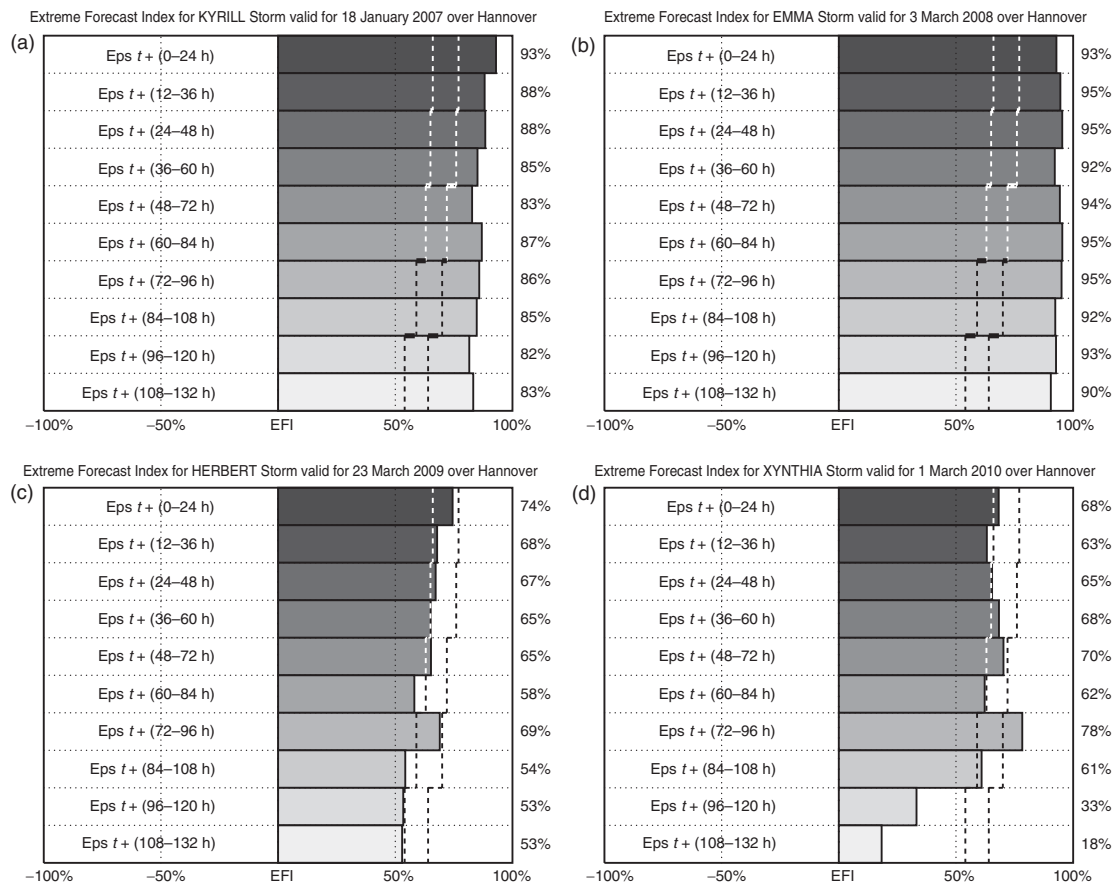


Figure 10. EFI-GRAMs for Hannover valid for: Kyrill (a), Emma (b), Herbert (c) and Xynthia (d) storms. The 95% (left) and 98% (right) EFI thresholds are highlighted by vertical dashed lines.

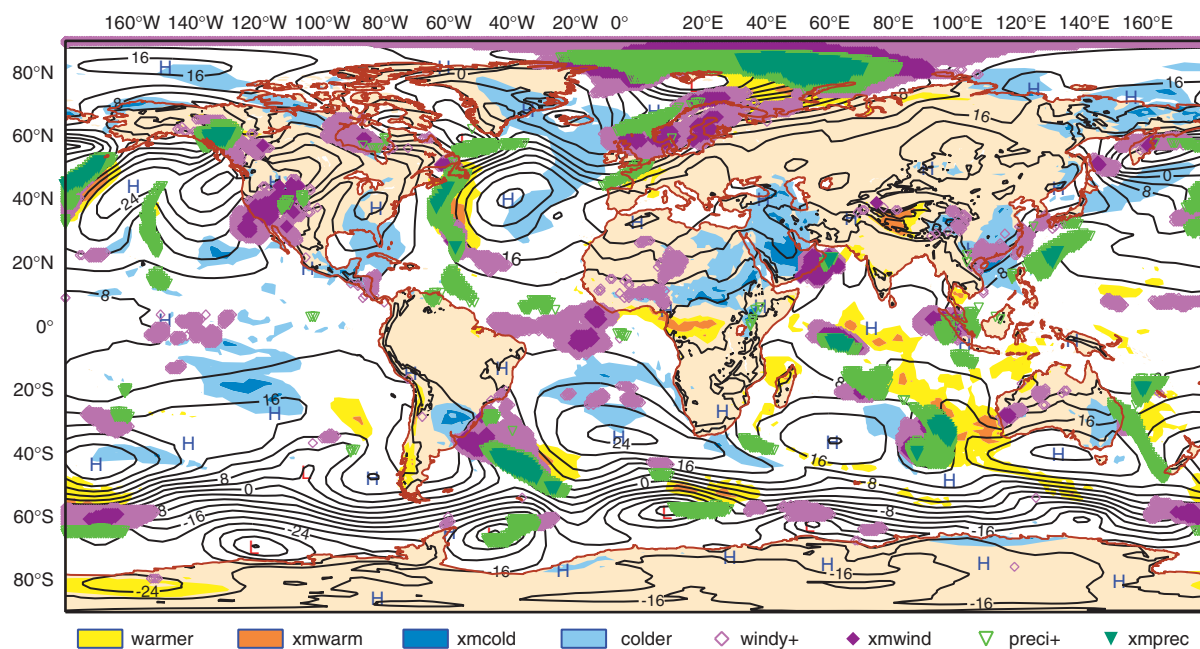


Figure 11. A composite I-EFI (Interactive-EFI) example valid for 1 December 2011. Purple symbols correspond to possible wind extremes; green symbols to rainfall extremes, while the rest of coloured symbols to possible temperature extremes. The EPS ensemble mean of 1000 hPa geopotential height is also plotted.

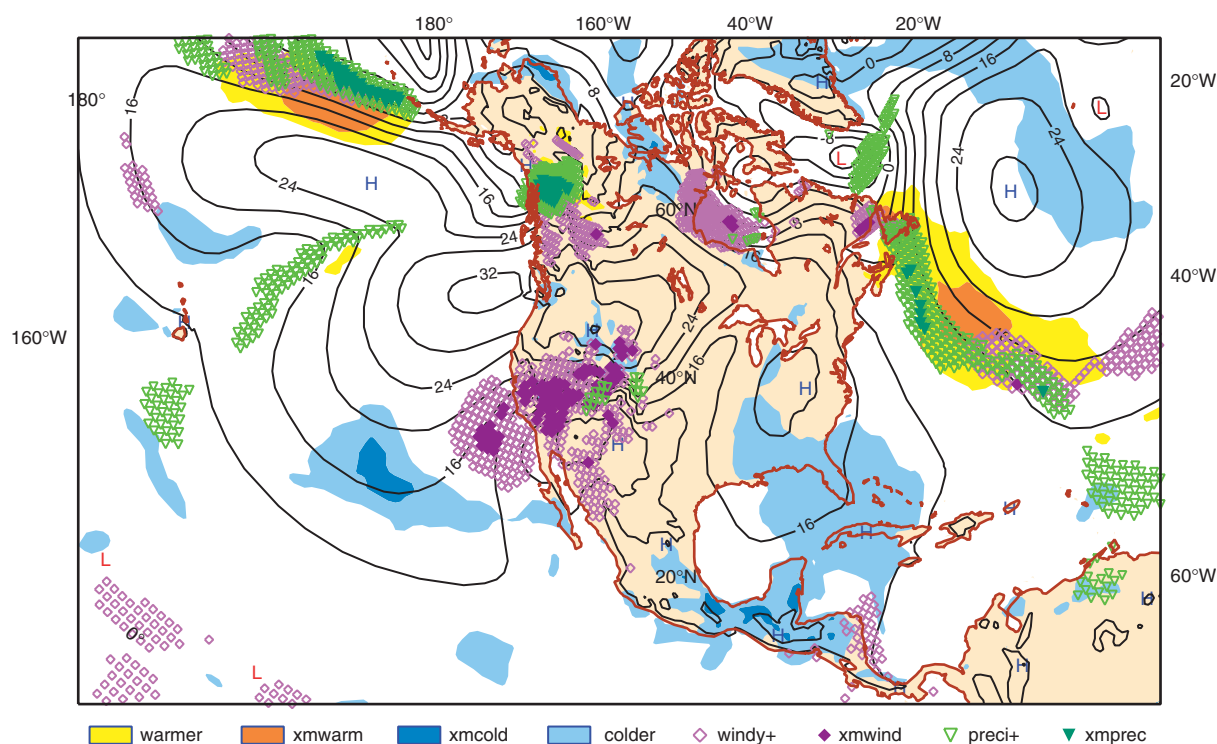


Figure 12. As in Figure 11, but zooming over North America.

weather predicted by ECMWF EPS for the 24 h interval of 1 December 2011, where different symbols correspond to different phenomena and intensity. Coloured symbols show the EFI warnings of extreme winds (purple symbols), heavy rainfall (green symbols) while the rest of the coloured symbols are used for expected (possible) temperature extremes.

By clicking the I-EFI map at any point, the user can display the CDF (Cumulative Density Function) distributions of EFI

values for the closest grid point. The location of Ogden Peak in northern Utah (41.28°N , 111.95°W) is used as an example, resulting to the set of graphs contained in Figure 13. Ogden Peak was hit really hard by the unusual extreme winds with an observed maximum of 91 mph (40.7 m s^{-1}). The CFDs for precipitation are contained in Figure 13(a), the corresponding 10 m wind gust in Figure 13(b), while the 2 m temperature is shown in Figure 13(c), valid for forecast horizons from 24

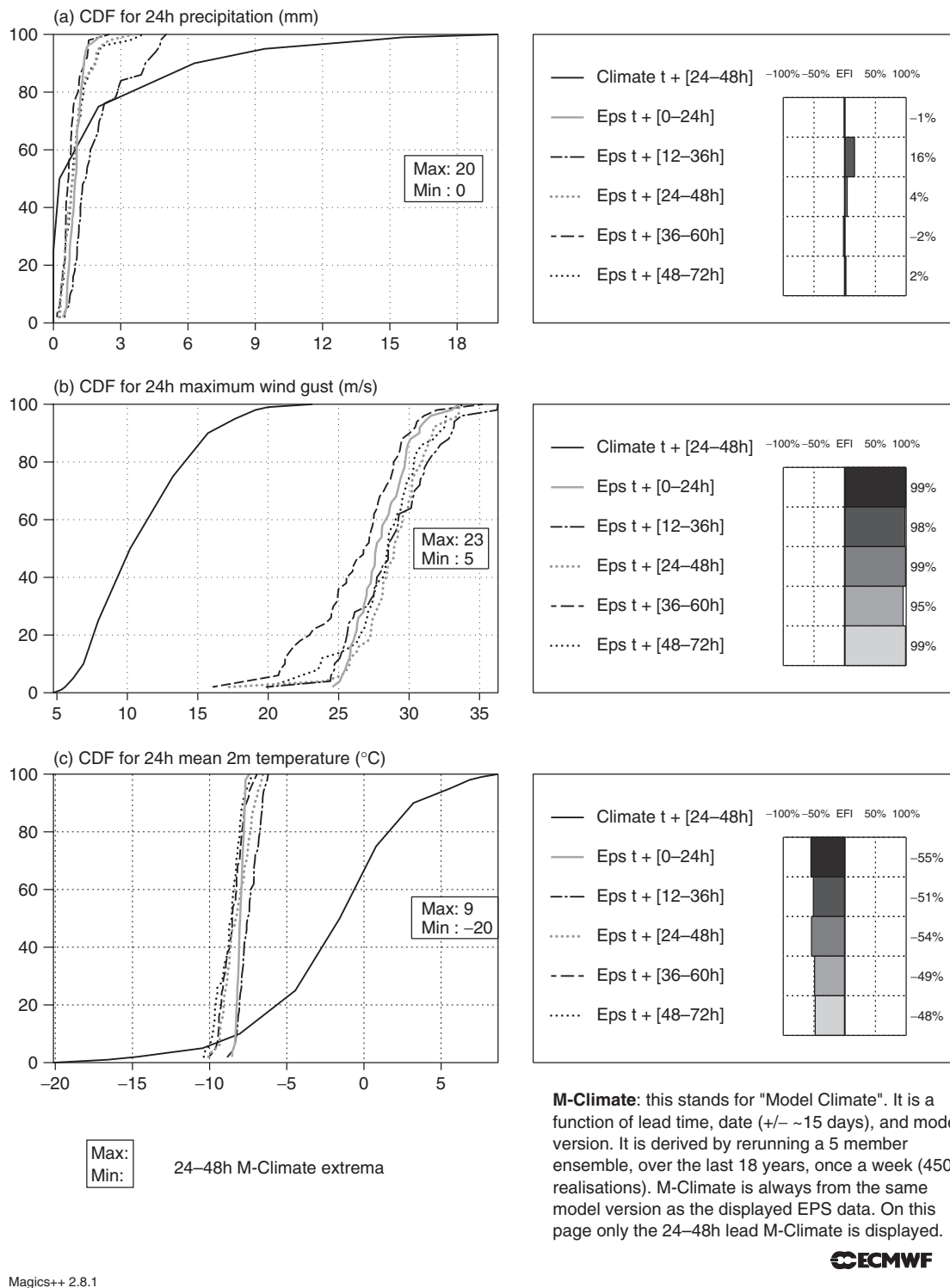


Figure 13. Cumulative Density Function (CDF) of EFI values for precipitation, wind gust and 2 m temperature over Ogden Peak, Utah, for 1 December 2011.

to 72 h in 12 h intervals. The maximum forecast horizon for such a product displayed in Figure 13 is 5.5 days. For the Figure 13(b) (EFI values based on gusts) a set of 10 EFI 10FGI maps as the ones contained in Figure 14 has to be used. Additional information coming from the EPS can be found in EPS-GRAMs such as the one presented in Figure 15 for Ogden Peak. The main ingredients (weather parameters) of such a composite diagram are:

- total cloud cover, which represents the instantaneous forecast value in oktas;
- total precipitation, which is the accumulated precipitation (sum of convective and large-scale precipitation);
- 10 m wind speed, which represents the instantaneous forecast value over the selected grid point, and,
- 2 m temperature, which is the instantaneous forecast value at 6 h intervals.

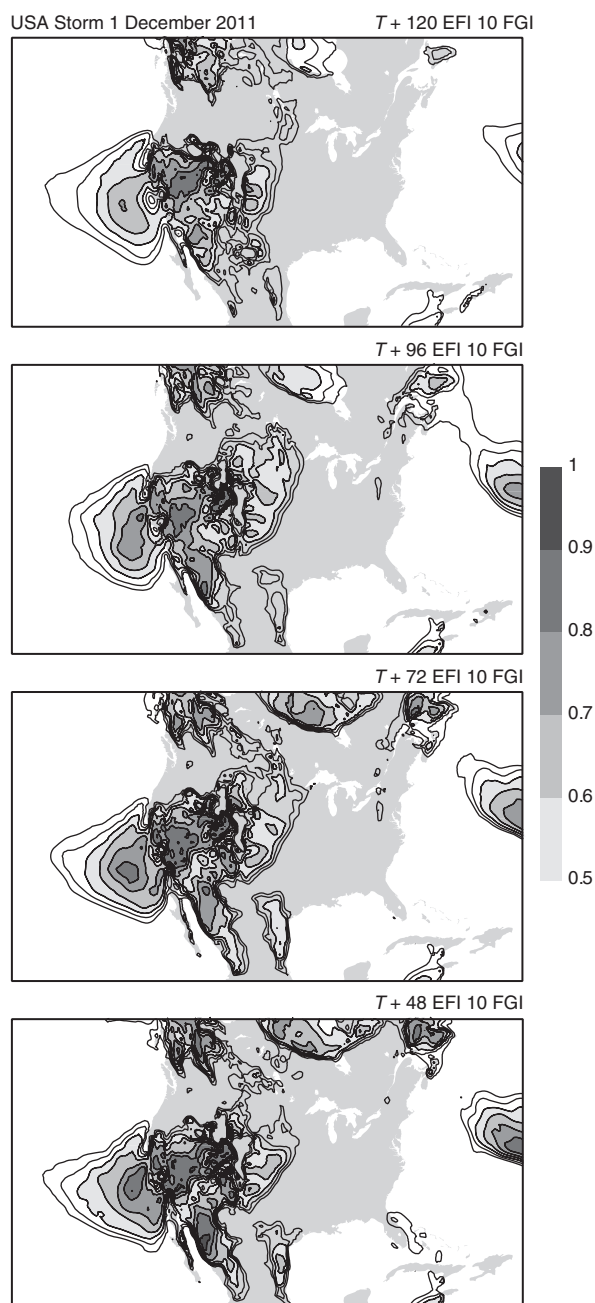


Figure 14. A sample of EFI maps being used in operational mode for the production of the middle panel of the 'EFI-GRAM' contained in Figure 13. Forecast horizons span from $T + 48$ to $T + 120$ h.

9. Overview

Extreme or anomalous events can be of mainly three types: (1) large-scale cold outbreaks or heat waves lasting for 3 days or more, (2) intense synoptic-scale dynamic precipitation and hurricane force winds, and, (3) strong organized sub-synoptic convection ('squall lines'). The EPS with its current resolution is well equipped to forecast the first two types of anomalous events falling in the synoptic scale. For smaller-scale extreme events, such as heavy rainfall, strong winds and rapid changes in temperature, forecast confidence decreases from day 3 onwards (Perrson, 2011).

In this study the focus is not only on such events but also on events of smaller scales that cannot be simulated exactly by the model, but some signals might still be apparent. Nevertheless,

with increasing resolutions the quality is steadily improving. The ability for models to generate severe storms has improved in recent years, although the direct comparison from model wind speed over land with observations shows a large, negative, bias. Among the reasons why this occurs is that modellers are mainly concerned with having a good momentum budget when designing boundary layer representations (Lalaurette and Grijn, 2005). However, a step towards the post-processing of maximum wind-gust values based on both explicit model winds and the subgrid-scale representation of turbulent fluxes for both the deterministic IFS and EPS was taken in 2000 (Lalaurette, 2001), resulting in a better adequacy between model and observations.

This study is focused mainly on early warnings in the short- and early medium-range (alerting). The backbone system for issuing early warnings has been the Extreme Forecast Index (EFI). For the assessment of the quality of the EFI, three synoptic stations (airports) of North Germany, i.e. Bremen, Hamburg and Hannover, were considered. An investigation of synoptic weather type for each station revealed that all wind extremes (belonging to the >99% extreme category) were linked to intense cyclonic circulation systems (i.e. surface pressure lows) being advected mainly by southwest and northwest regime flows.

Overall, it became clear that the first indication of an extreme wind event could come from the ability of the deterministic IFS and EPS components to capture very intense cyclonic circulation systems (reflecting possible windstorms) in the medium- and even late medium-range. These 'indications' may be communicated as 'alarm bells/signals'. Furthermore, in such cases that anticipated intense surface cyclonic circulation systems are linked to southwest or northwest advection flows, special attention should be given in the form of additional (critical) information. Of course, extra detail concerning time or place can be added nearer the time, including the possibility of issuing 'early warnings' based on the EFI.

For the objective evaluation of early warnings a set of different EFI formulations was linked to daily maximum wind speeds (in both 'Reanalysis' and 'Observation' modes). From all applications of the EFI schemes, the highest skill in issuing early warnings is given by the EFI-10FGI formulation based on the 0000 UTC forecast cycle. Although the ROCA (Area under the ROC Curve) values are found to be very high, suggesting a skilful performance, in a real operational mode use of the 99th EFI percentile threshold for 24 h forecasts, for example, would provide early warning for a considerable number of >99% category extremes, but not for all (i.e. 9 hits out of a total 24). By lowering this threshold the number of hits is increased until all extremes are captured (reflecting zero misses), but by doing so the number of false alarms is increased significantly. Consequently, an optimal trade-off between hits and false alarms has to be made when setting various (critical) EFI thresholds.

Already the EFI is a key resource for helping forecasters provide warnings of severe weather events. To provide additional assistance, ECMWF has immediate plans to extend the EFI out to 7 days.

Acknowledgements

SafeWind (European Commissions FP7 Project, Grant Agreement no. 213740) is to be acknowledged for supporting this work.

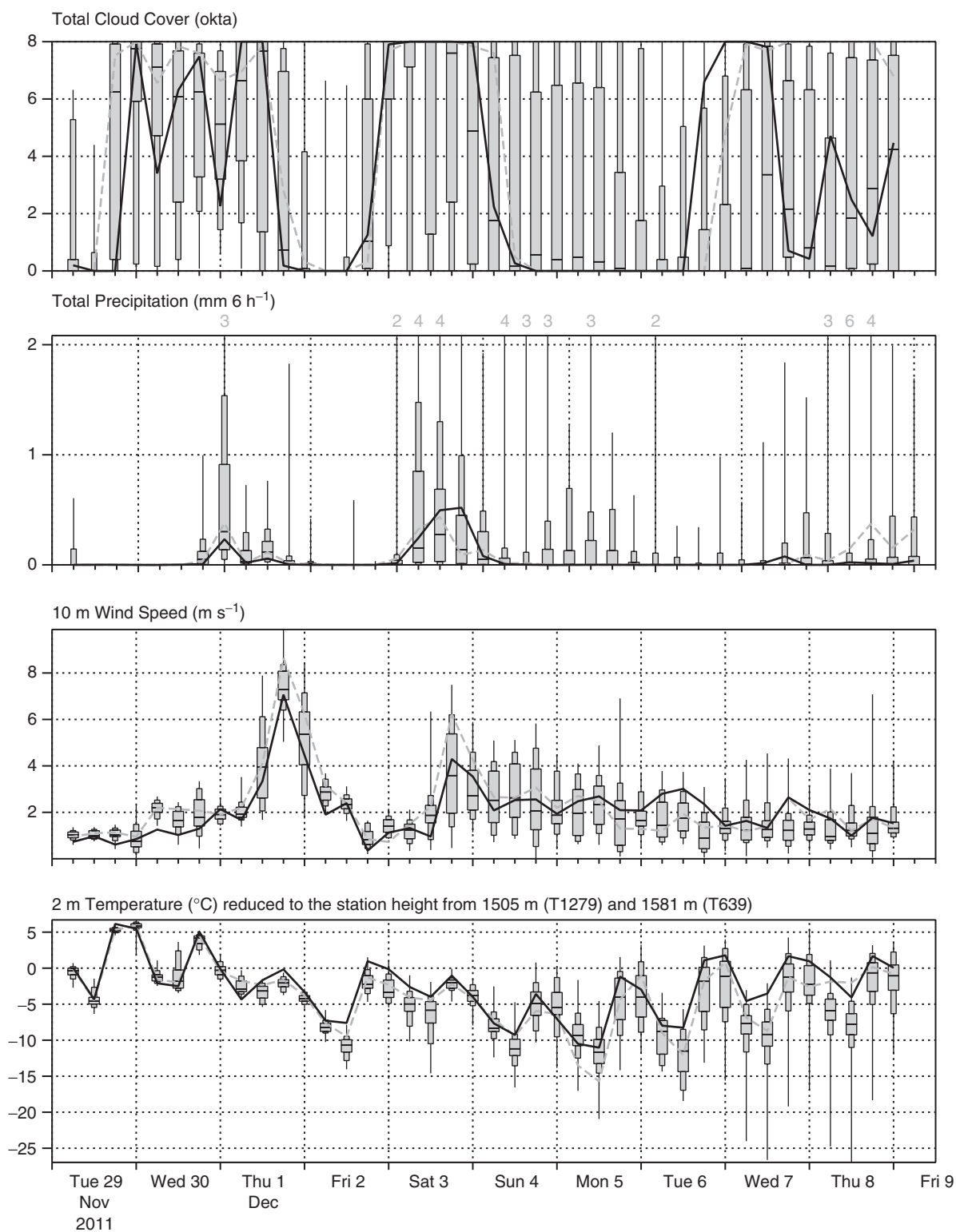


Figure 15. EPS-GRAM example corresponding to EFI-GRAM contained in Figure 13.

References

- Bissolli P, Dittmann E. 2001. The objective weather type classification of the German weather service and its possibilities of application to environmental and meteorological investigations. *Meteorol. Ztg.* **10**: 253–260.
- Bougeault P. 2003. *The WGNE Survey of Verification Methods for Numerical Prediction of Weather Elements and Severe Weather Events*. Meteo-France: Toulouse.
- Carpenter G. 2008. CAT-I catastrophe information reports, Windstorm Emma. <http://www.guycarp.com/portalapp/publicsite/catdocument.pdf?instratreportid=1677> (accessed 11 January 2012).
- Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A, Pocernich M, Damrath U, Ebert EE, Brown BG, Mason S. 2008. Review, forecast verification: current status and future directions. *Meteorol. Appl.* **15**: 3–18.
- Choo CW. 2009. Information use and early warning effectiveness: perspectives and prospects. *J. Am. Soc. Information Sci. Technol.* **60**: 1071–1082.

- Donat MG. 2010. European wind storms, related loss potentials and changes in multi-model climate simulations, Doktorarbeit am Institut für Meteorologie, Freie Universität, Berlin.
- Fink AH, Brucher T, Ermert V, Kruger A, Pinto JG. 2009. The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Nat. Hazards Earth Syst. Sci.* **9**: 405–423.
- Hess P, Brezowsky H. 1977. Katalog der Großwetterlagen Europas 1881–1976. 3. verbesserte und ergänzte. Aufl. *Ber. Dt. Wetterd.* **15**(113): Offenbach am Main.
- Lalurette F. 2001. Changes to the Operational Forecasting System. *ECMWF Newsletter* 89, Winter 1, 2000; Reading, UK.
- Lalurette F. 2003. Early detection of abnormal weather using a probabilistic Extreme Forecast Index. *Q. J. R. Meteorol. Soc.* **129**: 3037–3057.
- Lalurette F, van der Grijn G. 2005. Predictability of weather and climate: ensemble forecasts: can they provide useful early warnings?. In *Predictability of Weather and Climate*, Palmer TN, Hagedorn R (eds), Cambridge University Press: Cambridge, UK; 614–627.
- Legg TP, Mylne KR. 2004. Early warnings of severe weather from ensemble forecast information. *Weather Forecast.* **19**: 891–906.
- Meteo-France. 2010. The Xynthia windstorm of 27–28 February 2010, Meteo-France Actualités. http://france.meteofrance.com/france/actu/actu?portlet_id=50150&document_id=22089 (accessed 15 January 2012).
- Molteni F, Buizza R, Palmer TN, Petroliagis TI. 1996. The new ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**: 73–119.
- Murphy AH. 1991. Probabilities, odds and forecasts of rare events. *Weather Forecast.* **6**: 302–307.
- Nicolis C, Ebeling W, Baraldi C. 1997. Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes. *Tellus* **49A**: 108–118.
- Palmer TN, Buizza R, Leutbecher M, Hagedorn R, Jung T, Rodwell M, Vitart F, Berner J, Hagel E, Lawrence A, Pappenberger F, Park Y-Y, von Bremen L., Gilmour I. 2007. *The Ensemble Prediction System – Recent and Ongoing Developments*. Technical Memorandum 540. ECMWF: Reading.
- Palmer TN, Molteni F, Mureau R, Buizza R, Chapelet P, Tribbia J. 1993. Ensemble prediction. *Proceedings 1992 ECMWF Seminar: Validation of Models Over Europe*, ECMWF: Reading; 21–66.
- Persson A. 2011. User guide to ECMWF forecast products. <http://www.ecmwf.int/products/forecasts/guide> (accessed 11 January 2012).
- Richardson DS, Bidlot J, Ferranti L, Ghelli A, Haiden T, Hewson T, Janousek M, Prates F, Vitart F. 2011. Verification statistics and evaluations of ECMWF forecasts in 2010–2011.
- Simmons A, Burridge DM, Jarraud M, Wergen W. 1988. The ECMWF medium-range prediction models; Development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.* **40**: 28–60.
- Simmons A, Uppala S, Dee D, Kobayashi S. 2007. ERA-Interim: New ECMWF reanalysis products from 1989 Onwards. *ECMWF Newsletter No. 110*, 25–35.
- Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Academic Press: New York, NY.
- WMO. 2011. WMO brochure on extreme events 2001–2010, Weather extremes in a changing climate – Hindsight on foresight. <http://www.wmo.int/pages/mediacentre/news/documents/1075.en.pdf> (accessed 11 January 2012).